

STAT2911 Lecture Notes

Shane Leviton
6-6-2016

Contents

Table of discrete distributions	8
Introduction	9
Mathematical theory of probability.....	9
Sample space	9
Sample point	9
Examples:	9
Events:.....	9
Probability as a measure.....	10
Sigma algebra.....	10
Example of sigma algebra	11
Probability space:.....	11
Eg: model of fair dice	11
Venn diagrams:	11
De morgan's laws (for set theory):	12
Eg:.....	12
Equiprobable spaces	12
Claim: $PA = A\Omega$	13
Conditional probability:	14
Independence	15
Law of total probability.....	15
Distributive law for set theory:	16
Baye's rule:.....	16
Random Variables	16
Discrete RV.....	16
Probability mass function of Discrete RV:.....	16
Examples of random variables	17
Joint probability mass function:.....	20
Independence:	21
Poisson Distribution:.....	22
Assumptions of poisson:	22
Distibution of a sum of random variables: (arithmeticies of RV)	26
Convolution:.....	27
Expectation:	28
Definition	28
Examples of particular kinds	28

Bernoulli:	28
Poisson	28
Binomial	29
Geometric:	29
Negative binomial	29
Hypergeometric	30
Negative expectation:	30
Expectation of a function of a RV:	30
Proof:	30
Proof of 2:	31
Variance	31
Definition:	32
Variance and expectation computation	32
Examples of variance	32
Table of discrete distributions	34
Expectation as a linear operator	35
Expected value of sum of $X + Y$	35
Random vector	35
Expectation multiplication	37
Claim:	37
L1 and L2	38
Example of $X \in L1$ but not $\in L2$	38
What about the opposite?	38
So is $VX + Y \in L2$?	38
Claim:	38
Claim: shifting $V(X)$	39
Covariance	40
Definition	40
Claim:	40
So; going back to: $VX + Y$	41
By induction: variance of X_i 's	42
Chebyshev's inequality	44
Lemma: Markov's Inequality	45
Law of large numbers:	46
Convergence probability	46
Theorem: Weak law of large numbers	47

Strong law of large numbers:.....	47
The multinomial distribution:	47
Multinomial random vector:.....	48
Goal: pm for multinomial Random vector	48
Estimation:	49
The method of moments	49
Method of moments	49
For estimating	51
Maximum likelihood estimation (MLE).....	53
Likelihood function	53
Hardy-Weinberg equilibrium	59
theory.....	59
Example:.....	59
So how do we rigorously compare different estimators?	61
Back to HW:.....	63
Hardy – Weinberg equilibrium.....	65
Delta method:	66
Parametric Bootstrap method	72
Example: Binomial.....	72
Example 2: Hardy Weinberg equilibrium	73
Conditional Expectation/Variance	74
Conditional expectation:.....	74
Example: die.....	75
Claim: $L1$	75
Random Sums:	75
Examples:	75
Conditional Expectation (the RV).... (not a number, a random variable).....	77
Definition:	77
Theorem: Expectation of conditional expectation (total expectation law)	79
What about the variance of conditional expectation? $VEYX$ is it VY ?	80
Conditional Variance:.....	81
Analogously:.....	81
So, conditional variance:.....	81
Continuous Random Variables.....	83
CDF (Cumulative distribution function)	83
Definition:	83

Continuous random variable definition:.....	86
PDF: f	86
Examples of PDF/CDF distribution:.....	88
Uniform distribution:	88
Exponential distribution.....	89
Gamma distribution:	91
Normal distribution.....	93
Quantiles:.....	94
Quantile definition:.....	94
Special quantiles	95
Pth quantile.....	95
pth quantile definition:	95
Quantile function: Definition	96
Functions of random variables for continuous RV.....	97
Claim: uniform distribution.....	97
Claim: sampling	98
SAMPLING!!!.....	99
Claim:	99
Theorem: Random variable functions	99
Examples of samples and functions:.....	100
Joint distribution (discrete analogue of joint pmf)	102
Joint density definition:	103
Comments:.....	103
Marginal distribution:	105
Proof:	105
Examples:	106
Gamma/exponential joint.....	106
Uniform distribution of region in a plane	107
Bivariate normal (special case)	109
Independent RV's (continuous)	110
For continuous:	110
Definition:	110
Examples:	111
Conditional distributions:	113
Continuous:.....	113
Construction:.....	113

Conditional density:	114
Examples:	115
Sampling in 2D	116
Discrete:	116
Continuous	116
Sum of Random Variables	116
Continuous:	116
Density of Sum of continuous random variables $Z = X + Y$: (convolution)	117
Quotients of RV's	119
Formula for quotients:	122
For independent RV	122
Functions of jointly continuous RV	122
Mapping:	122
Blow up factor:	123
Jacobian: generally	124
Box-Muller Sampling	128
Visual comparison of distributions	129
QQ plot	129
Examples:	129
More general QQ plot:	130
Empirical data:	130
Extrema and order statistics	137
Sorted values: order statistics	137
Claim:	137
Expectation of Order statistics	139
Expectation of a continuous RV	139
Definition:	139
Looking at examples:	139
Expectation of a function for continuous RV	142
In continuous case analogue: Expectation of function	142
Variance for continuous:	143
Equation to calculate:	144
Variance of linear sum:	144
Covariance:	144
Claim:	144
Variance of sum of RVs independent:	144

Standard deviation:.....	144
Correlation coefficient:	145
Claim: Correlation coefficient	145
Examples of variance ect:	145
Standard normal:	145
General normal: $X \sim N\mu, \sigma^2$	146
Gamma.....	146
Bivariate standard normal:	147
Markov's inequality in Continuous:	148
continuous	148
Chebyshev's inequality:.....	148
Weak law of large numbers:	149
Strong law of large numbers:.....	149
Estimation for continuous.....	149
Examples:	149
Standard normal:	149
Normal:	149
Exponential:	149
Scaled Cauchy:	149
Method of moments:.....	149
Examples:	150
Maximum likelihood estimation (MLE).....	150
For continuous:	151
Conditional expectation:.....	152
Conditional expectation def:	152
More generally:.....	152
Mixed distribution.....	153
Examples	154
Uniform and binomial	154
Natural valued with iid sum	154
Expectation of mixed distributions:	154
Eg 1:.....	154
Eg 2:.....	155
Variance of random sum:.....	155
Law of total probability	155
We can now prove this:	155

Prediction	156
Predictor	156
Claim:	157
Back to prediction problem:	157
Example of prediction:	158
Moment generating function (MGF)	160
Definition:	160
Calculating MGF:	160
So why are we doing this?	161
Theorem 1:	161
Theorem 2:	163
Weak convergence/converges in distribution	164
Weakly converges	164
Characteristic function (CF)	167
Confidence intervals	168
Estimation:	168
We want to know- how close are we to θ ?	168
Confidence interval set up:	169
Constructing confidence intervals (CI s)	172
Examples of confidence intervals:	172
Constructing approximate CIs based on the MLE	180
Confidence intervals for Bernoulli θ :	181

PROBABILITY AND STATISTICAL MODELS

STAT2911 Notes

Table of discrete distributions

Distribution	Model	pmf	$E(X)$	$V(X)$
<i>Bernoulli</i> (p)	Success or failure, with probability p Eg: coin toss	p	p	$p(1-p)$
<i>Geometric</i> (p)	Probability p to first success: eg coin toss till first head	$(1-p)^{k-1}p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
<i>Binomial</i> (n, p)	n Bernoulli trials. Eg n coin tosses	$\binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$
<i>Poisson</i> (λ)	Number of events in a certain time interval. Eg: number of radioactive particles emitted in certain time	$e^{-\lambda} \frac{\lambda^k}{k!}$	λ	λ
<i>Hypergeometric</i> (r, n, m)	number of red balls r in a sample of m balls drawn without replacement from an urn with r red balls and $n-r$ black balls	$\frac{\binom{r}{k} \binom{n-r}{m-k}}{\binom{n}{m}}$	$\frac{mr}{n}$	$m \frac{r}{n} \frac{n-r}{n} \frac{n-m}{n-1}$
<i>Negative binomial</i> (r, p)	Number of Bernoulli trials till the r^{th} success	$\binom{k-1}{r-1} (1-p)^{k-r} p^r$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$

Uri Keich; Carlslaw 821; Monday 5-6

Introduction

- Probability in general and this course has 2 components:
 - o Mathematical theory of probability
 - Definitions, theorems and proofs
 - Abstraction of experiments whose outcome is random
 - o Modelling: argued rather than proved applications can be confusing, as well as ill defined
 - Useful for describing/summarising the data and making accurate predictions

Mathematical theory of probability

Sample space

The set of all possible outcomes, denoted Ω , is the sample space

Sample point

A point $\omega \in \Omega$ is a sample point

Examples:

Die:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Coin:

$$\Omega = \{H, T\}$$

- Complication: do we model the possibility that the coin can land on its side?

Events:

- Events are subsets of Ω , for which we can assign a probability.
- We say an event A occurred if the outcome, or sample point $\omega \in \Omega$ satisfies $\omega \in A$

Probability as a measure

$P(A)$ is the probability of the event A , which **intuitively** is the rate at which A occurs if we repeat the experiment many times.

Mathematically: P is a probability measure function if:

1. $P(\Omega) = 1$
2. $P(A) \geq 0$ for any event A
3. If $A_1, A_2, \dots$ are mutually disjoint ($A_1 \cap A_2 \cap \dots = \emptyset$) then $(P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n))$

For this to make sense, we need to know that a union of a sequence of events is ALSO an event.

- Why do we bother with determining events? Why can't any subset of Ω be an event?
 - o Imagine choosing a point at random in a large cube $C \subset \mathbb{R}^3$, where the probability of the point lying in any set $A \subset C$ is proportional to its volume $|A|$
 - o Clearly, if $A, B \in C$ are related through a rigid motion (translation + rotation), then $|A| = |B|$ so their probabilities are the same
 - o Similarly, if $A \in C$ can be split into $A_1 \cup A_2$, then $P(A) = P(A_1) + P(A_2)$
 - o Bernard-Tarski Paradox:
 - The unit ball B in $\mathbb{R}^3$ can be decomposed into 5 pieces, which can be assembled using only rigid motions into two balls of the same size.
 - But then: $P(B) = P(B_1) + P(B_2) + \dots + P(B_5) = P(B'_1) + \dots + P(B'_5) = 2P(B)$
 - This is why we cannot assign probabilities to EVERY subset of Ω . Probabilities can't be assigned to arbitrary sets, rather to sets which are "measurable".

The collection of measurable sets is captured by the notion of σ algebra (σ -field)

Sigma algebra

Definition:

A collection of subsets of Ω is a σ - algebra is

1. $\Omega \in F$
2. $A \in F \Rightarrow A^c \in F$
3. $A_1, A_2, \dots \in F \Rightarrow \cup_{n=1}^{\infty} A_n \in F$

2. says that F is closed with respect to complementing and taking the complement

3. says that F is closed with respect to a countable union

- (countable means you can count it)

$A = \{1,3,5\}$ is a finite countable set

- $\mathbb{N}$ is an infinite countable set
- $\mathbb{Z}$ is also an infinite countable set
- $\mathbb{Q}$ is infinite countable
- $\mathbb{R}$ is not countable

Example of sigma algebra

$$\Omega = \{1, 2, \dots, 6\}$$

F = the power set of Ω = the set of all subsets of Ω

$$= \{\emptyset, \{\text{all singles}\}, \{\text{all doubles}\}, \dots, \Omega\}$$

Denoted

$$= 2^\Omega$$

Questions:

What is the cardinality of F , or how many subsets of Ω does it contain

-

Is 2^Ω ALWAYS a σ algebra

What is the smallest σ algebra regardless of the sample space

ANSWERS?

- $2^{|\Omega|}$?

- $\{\emptyset, \Omega\}$

Probability space:

A probability space consists of

1. A sample space Ω
2. A σ – algebra of subsets of Ω , F
3. A probability measure: $P: F \rightarrow \mathbb{R}$

Eg: model of fair dice

$$\Omega = \{1, 2, \dots, 6\}$$

$$F = 2^\Omega$$

$$P(\{i\}) = \frac{1}{6}, \quad i = [1, 6]$$

$$P(A) = \frac{|A|}{6} \text{ (for } A \in F)$$

Venn diagrams:

Useful to visualise relations between sets. They help plan proofs, but don't use them as PROOFS in a course

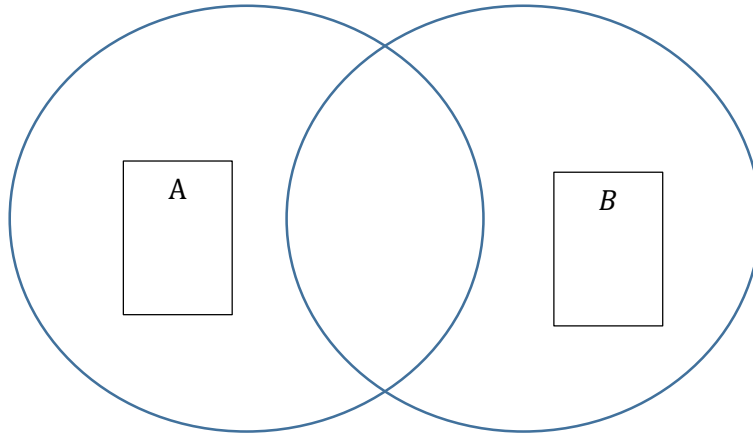
De Morgan's laws (for set theory):

$$(A \cup B)^c = A^c \cap B^c$$

$$(A \cap B)^c = A^c \cup B^c$$

Eg:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



- Why is $A \cup B$ and $A \cap B \in \mathcal{F}$? (why are they events)

$$A \cup B = A \dot{\cup} (B \setminus A) \quad [(\dot{\cup}) = \text{disjoint union}]$$

$$B \setminus A = B \cap A^c$$

$$\therefore P(A \cup B) = P(A) + P(B \setminus A)$$

$$\left(\text{using that } P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n), \quad \text{with } A \text{ and } B \text{ and infinite } \emptyset \text{'s} \right)$$

(probability of empty set is 0)

$$P(\emptyset)P(\bigcup \emptyset) = \sum_{n=1}^{\infty} P(\emptyset) \rightarrow P(\emptyset) = 0$$

$$P(B) = P(B \setminus A) + P(B \cap A)$$

$$\rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Equiprobable spaces

- A space consists of a finite sample space such that $\forall \omega \in \Omega \quad P(\{\omega\}) = c > 0$
- (all equally likely)

Note: $\omega \in \Omega$ is a sample point, $\{\omega\}$ = an event in F

- Why does it have to be finite?
 - o Otherwise

$$\begin{aligned}\{\omega\}_{n=1}^{\infty} &\subset \Omega \\ P(\cup_n \{\omega_n\}) &= \sum_n P(\omega_n) = \sum_{n=1}^{\infty} c = \infty \\ [but \forall A \subset F, P(A) \in [0,1] \text{ (why?} \rightarrow \text{show)}]\end{aligned}$$

Claim: $P(A) = \frac{|A|}{|\Omega|}$

Proof:

$$\begin{aligned}A &= \cup_{\omega \in A} \{\omega\} \\ \therefore P(A) &= P(\cup_{\omega \in A} \{\omega\}) = \sum_{\omega \in A} P(\omega) \\ &= |A| \times c \\ \text{take } A &= \Omega \\ \rightarrow P(\Omega) &= 1 \text{ (definition)} = |\Omega|c \\ \rightarrow c &= \frac{1}{|\Omega|} \\ \rightarrow P(A) &= \frac{|A|}{|\Omega|}\end{aligned}$$

This means: that probability is just combinatorics!!!!

Example of probability combinatorics

1. A fair die is rolled:

$$P(\{\omega\}) = \frac{1}{6}$$

2. A group of n people meet at a party, what is the probability that at least 2 of them share a birthday
 - o model that no leap years, no association between birthdays

$$\begin{aligned}\Omega &= \{(i_1, \dots, i_n) \mid i_k \in \{1, \dots, 365\}\} \\ \rightarrow \frac{1}{|\Omega|} &= 365^{-n} \\ A &= \{(i_1, \dots, i_n) \in \Omega \mid \exists j \neq k, i_j = i_k\} \\ \text{Let's rather compute the complement} \\ A^c &= \{(i_1, \dots, i_n) \in \Omega \mid \exists j \neq k, i_j \neq i_k\}\end{aligned}$$

$$\begin{aligned}
 P(A^c) &= \frac{|A^c|}{|\Omega|} = \frac{365 \times 364 \times \dots (365 - n + 1)}{365^n} \\
 &= \prod_{i=1}^n \frac{365 - (i - 1)}{365} \quad \{n \leq 365\} \\
 &= \prod_{i=1}^n 1 - \frac{i - 1}{365} \\
 P(A) &= 1 - \frac{\binom{365}{n}}{365^n}
 \end{aligned}$$

Sidenote: for $n = 23 \rightarrow P(A) \approx \frac{1}{2}$

3. what is the probability that one of the guests shares YOUR birthday

$$B = \{(i_1, \dots, i_n) \in \Omega \mid \exists j, i_j = x \text{ (your birthday)}\}$$

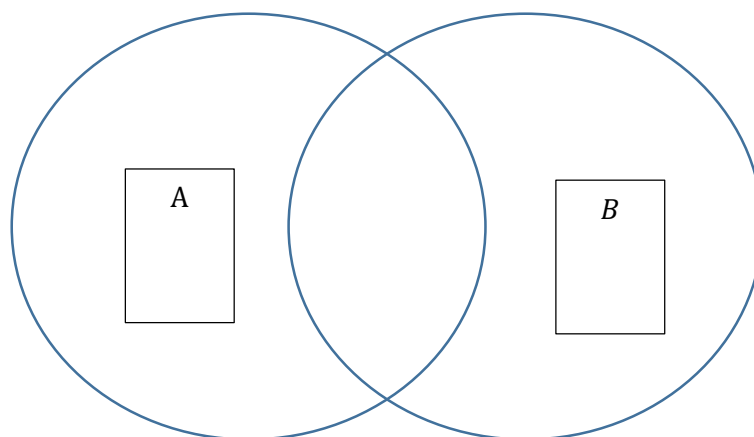
$$\begin{aligned}
 \therefore B^c &= \{(i_1, \dots, i_n) \in \Omega \mid \forall j, i_j \neq x\} \\
 &= |B^c| = 364^n \\
 \rightarrow P(B^c) &= \frac{364^n}{365^n} \\
 &= \left(1 - \frac{1}{365}\right)^n \approx 3^{-\frac{n}{365}}
 \end{aligned}$$

Sidenote: for $n = 253, P(B) \approx \frac{1}{2}$

Conditional probability:

If $A, B \in \mathcal{F}$ and $P(B) > 0$ we can define (Probability of A given B)

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$



Example:

$$\Omega = \{1, \dots, 6\}$$

$$A = \{1,2\}; B = \{1,3,5\}$$

Independence

Event $A \in F$ is independent (ind) of $B \in F$ if knowing whether or not A occurred, does **not** give us any information on whether or not B occurred.

$$\therefore P(B|A) = P(B)$$

$$\rightarrow \frac{P(B \cap A)}{P(A)} = P(B)$$

$$\therefore P(B \cap A) = P(A)P(B)$$

- this definition is more robust (symmetric about A and B), and no need to assume $P(A) > 0$; and easier to generalise

general:

$$P(A_1 \cap A_2, \dots \cap A_n) = P\left(\prod_{i=1}^n A_i\right)$$

- stronger than pairwise independence!!!

NOTE: independence is NOT DISJOINT (disjoint gives you total knowledge that the other would not have occurred)

Law of total probability

- a probability of an event may be computed by summing over all eventualities

3 machines	A	B	C
Production rate	.05	.2	.3
Failure rate	.01	.02	.005

$$P(\text{failed product}) = .01(.05) + .2(.02) + .3(.005)$$

Generally:

If $P_j \in F$ forms a partition of Ω

$$\Omega = \dot{\cup}_j B_j$$

$$\text{then } P(A) = P(A \cap \Omega)$$

$$= P\left(A \cap \left(\dot{\cup}_j B_j\right)\right) = P\left(\dot{\cup}_j (A \cap B_j)\right) = \sum_j P(A \cap B_j) = \sum_j P(A)P(B_j|A)$$

Distributive law for set theory:

$$\begin{aligned}A \cup (B \cap C) &= (A \cup B) \cap (A \cup C) \\A \cap (B \cup C) &= (A \cap B) \cup (A \cap C)\end{aligned}$$

Baye's rule:

Diagnostic: which of the events B_j triggered the event A ?

$$P(B_j|A) = \frac{P(A \cap B_j)}{P(A)}$$

(which machine B caused the failure of A?)

$$= \frac{P(A|B_j)P(B_j)}{\sum_i P(A|B_i)P(B_i)}$$

(using definitions of law of total probability and conditional probability)

Random Variables

A RV is a measurable function $X: \Omega \rightarrow \mathbb{R}$

Discrete RV

A RV is discrete if its range:

$$X(\Omega) = \{X(\omega): \omega \in \Omega\} \subset \mathbb{R}$$

Is a countable set (finite or infinite)

Probability mass function of Discrete RV:

The pmf of X is defined as:

$$p_X(x) = P(X = x) = P(\{\omega: X(\omega) = x\}) \quad [\text{with } x \in \mathbb{R}]$$

- why is $\{\omega: X(\omega) = x\} \in \mathcal{F}$?

Properties of pmf

Claim:

1. $\forall x \in \mathbb{R} \setminus X(\Omega)$ then $p_X(x) = 0$ (the pmf of something which is unattainable is 0)
2. With $\{x_i\} = X(\Omega)$, $\sum_i p_X(x_i) = 1$ (the pmf of all outcomes is 1)

The distribution of a discrete RV is completely specified by its pmf. Indeed, $\forall A \subset \mathbb{R}$

$$P(X \in A) = \sum_{i: x_i \in A} p_X(x_i)$$

(shows the probability that an outcome is going to happen)

- We can thus specify the distribution of a RV X by specifying its pmf p_X

Question:

Can any functions $p: \mathbb{R} \rightarrow [0,1]$ with a countable support $\{x: p(x) > 0\}$ such that $\sum_{i: p(x_i) > 0} p(x_i) = 1$ be a pmf for some random variable?

Claim: If p is as above, then there exist a probability space (Ω, F, p) about a RV $X: \Omega \rightarrow \mathbb{R}$ such that $p_X = p$

Proof: $\Omega = \{x: p(x) > 0\}; F = 2^\Omega; P(A) = \sum_{x \in A} p(x)$

$$X(\omega) = \omega$$

Examples of random variables

Bernoulli random variable

Is defined by the pmf

$$p(x) = \begin{cases} 1-p; & x=0 \\ p; & x=1 \end{cases} \quad (\text{failure and success})$$

Eg:

$$\begin{aligned} \Omega &= \{H, T\}; \quad P(H) = p \\ X(\omega) &= \begin{cases} 1; & \omega = H \\ 0; & \omega = T \end{cases} \end{aligned}$$

Binomial random variable

S_n models the number of successes in an iid (independent and identically distributed) Bernoulli trials

A 2 parameter family of distributions: (n, p)

Note: if $X_i = 1$: {success of trial} = $\begin{cases} 0 & \text{if failure} \\ 1 & \text{if success} \end{cases}$; then $X_i \sim \text{Bernoulli}(p)$ and $S_n = \sum_{i=1}^n X_i$

Generally: for an event A , then the random variable 1_A is the indicator function of A :

$$1_A(RV) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \omega \text{ not in } A \end{cases}$$

$$S_n(\Omega) = \{0, 1, 2, \dots, n\}$$

For $k \in S_n(\Omega)$; $p(S_n = k)$ is:

- Consider the configuration of a Bernoulli trials: $s, s, s, \dots, f, f, f, \dots$
- Its probability is $p^k (1-p)^{n-k} \times \binom{n}{k}$

Binomial RV pmf

$$\therefore p(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Geometric random variable:

Models the number of iid Bernoulli (p) trials if it takes till the first success

$$X(\Omega) = \{1, 2, 3, \dots\} \cup \{\infty\}$$

A 1-parameter family : p

Pmf of geometric random variable

$$p(X = k) = (1 - p)^{k-1}p$$

$$P(X > k) = (1 - p)^k$$

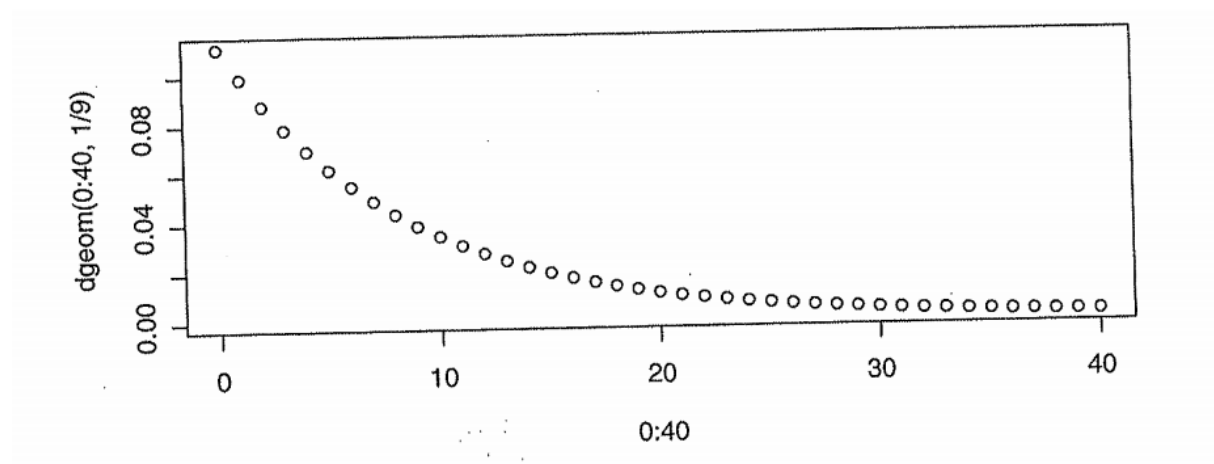
$$P(X > n + k | X > k) = \frac{P(X > n + k, X > k)}{P(X > k)} \quad (\text{using conditional probability})$$

$$= \frac{P(X > n + k)}{P(X > k)} = \frac{(1 - p)^{n+k}}{(1 - p)^k} = (1 - p)^n = P(X > n)$$

This property is “memoryless” (the probability does not remember what has happened before)

- Geometric is the ONLY discrete memoryless distribution

If $x_i = 1_{\{\text{success of trial } i\}}$, then $x = \min\{i: X_i = 1\}$



Negative binomial RV

Models the number of iid Bernoulli trials till the r^{th} success, where $r \in \mathbb{N}$

- A 2 parameter distribution
- Note: $X^1: r = 1$ is a geometric random variable

$$X^r(\Omega) = \{r, r + 1, \dots\} \cup \{\infty\}$$

Pmf:

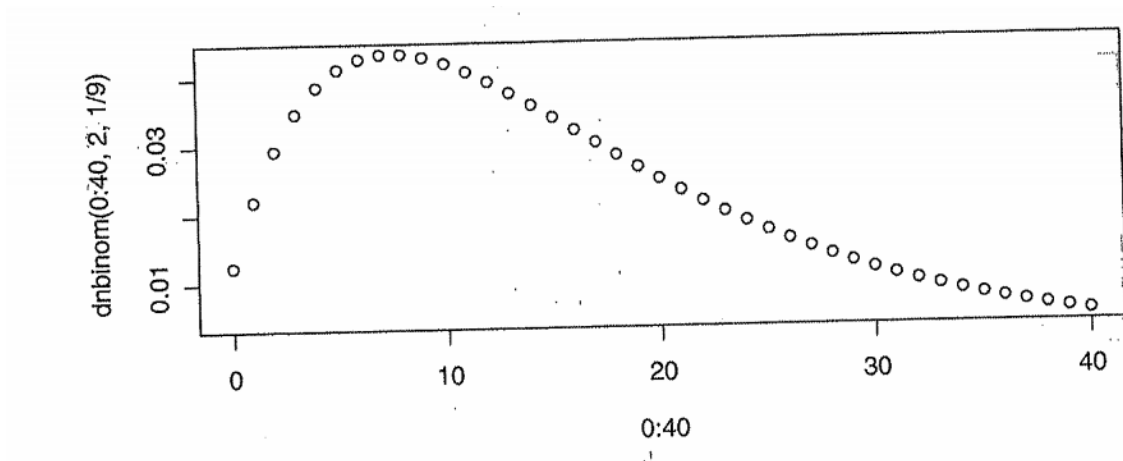
For $k \in X^r(\Omega)$, $P(X^r = k) =$

- Consider the configuration of its Bernoulli trials

$$(f, f, f, \dots, s, s, s)$$

$$p = \binom{k-1}{r-1} (1-p)^{k-r} p^r$$

$$X^r = \min \left\{ m: \sum_{i=1}^n x_i = r \right\}$$



Hypergeometric RV

X models the number of red balls in a sample of m balls drawn **without** replacement from an urn with r red balls and $n - r$ black balls

$$X(\Omega) \subset \{0, 1, \dots, r\}$$

Probability:

For $k \in X(\Omega)$

$$pmf: P(X = k) = \frac{\binom{r}{k} \binom{n-r}{m-k}}{\binom{n}{m}}$$

Which is a 3 parameter family: (r, n, m)

(is not phrased in iid Bernoulli RV)

- A famous case of this is the Fisher Exact Test

Fisher Exact Test:

30 convicted criminals with same sex twin of 13 which 13 were identical and 17 were non identical twins.

- Is there evidence of a genetic link?

	Convicted	Not convicted	
Identical	10	3	13
different	2	15	17
	12	18	

- Assuming that whether or not the twin of the criminal is also convicted does not depend on the biological type of the twin, we have a sample from a hypergeometric distribution

Indeed: there are 13 red (monozygote) balls and 17 black (dizygotic) balls. We randomly sample 12 balls (convicted), what is the probability that we will see 10 or more red balls in the same sample?

Let $X \sim \text{Hyper}(13, 17, 12) = (r, k, m)$

$$P(X \geq 10) = \sum_{k=0}^{12} \frac{\binom{13}{k} \binom{17}{12-k}}{\binom{30}{12}} \approx 0.000465$$

So- it seems very unlikely that there is no relation between the conviction of the twin and its type, but:

- Did we establish that a criminal mind is inherited?
 - o Problem with this statement
 - o Conviction vs truth "that face is up to no good??"
 - o Sampling/ascertainment bias
 - o Identical twins tighter connection?

Joint probability mass function:

The joint pmf of the RVs X and Y specifies their interaction:

$$P_{XY}(x, y) = P(X = x, Y = y) \quad (x, y \in \mathbb{R})$$

Note: $\{X = x, Y = y\}; \{\omega \in \Omega \mid X(\omega) = x, Y(\omega) = y\}$

- If X and Y are discrete Random Variables, with
 - o $\{x_i\} = X(\Omega)$; and $\{y_i\} = Y(\Omega)$

Then: (keeping y fixed)

$$P_X(x) = P(X = x) = \sum_j P(X = x, Y = y_j) = \sum_j P_{XY}(x, y_j)$$

Similarly:

$$P_Y(y) = \sum_i P_{XY}(x_i, y)$$

P_X and P_Y are referred to as the **marginal pmf's**

Example:

A fair coin is tossed 3 times

$X = 1 \{H \text{ on first toss}\} = \text{number of heads in first toss}$

$Y = \text{total number of heads}$

$$\Omega = \{HHH, HHT \dots\}; |\Omega| = 8$$

Y	0	1	2	3	
X	1/8	2/8	1/8	0	1/2
0	0	1/8	2/8	1/8	1/2
1	0	1/8	2/8	1/8	
	1/8	3/8	3/8	1/8	

So, $P(X = 0) = \frac{1}{8}$

(note: called marginal, as it is in the margin)

NOTE:

$$X \sim \text{Bernoulli} \left(\frac{1}{2} \right)$$

$$Y \sim \text{Binomial} \left(3, \frac{1}{2} \right)$$

Are X and Y independent RVs?

- The random variable X is independent of Y if “knowing the value of Y does not change the distribution of X ” (so NO- they are not independent)

Independence:

$$P(X = x_i, Y = y_j) = P(X = x_i, Y \neq y_j)$$

$$\rightarrow P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j) \text{ (Definition)}$$

$$P_{XY}(x, y) = P_X(x)P_Y(y)$$

The joint pmf factors into the product of the marginal

More generally, the random variables $X_1, \dots, X_N$ are independent if:

$$P(X_1 = x_1, \dots, X_n = x_n) = P_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_i^n P_{X_i}(x_i) \quad (\forall x_i \in \mathbb{R})$$

Note: take caution that PAIRWISE INDEPENDENCE DOES NOT IMPLY INDEPENDENCE

Poisson Distribution:

Recall:

$$X \sim \text{Poisson}(\lambda)$$

If

$$P_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

$$(k \in \mathbb{N})$$

- A 1 parameter family, $\lambda > 0$
- Where does it come from?
 - o Models the number of “events” that are registered in a certain time interval, eg: the number of
 - Particles emitted by a radioactive source in a hour
 - Incoming calls to a service centre between 1-2 pm
 - Light bulbs burnt in a year
 - Fatalities from horse kicks in the Prussian cavalry over 200 corp years (Boriewicz 1898)

20 corps x 10 years = 200 corp years

Number of deaths	Observed count	frequency	Poisson approximation
0	109	.545	.543
1	65	.325	.331
2	22	.110	.101
3	3	.015	.021
4	1	.005	.003

So, how did we come up with the Poisson distribution?

Assumptions of poisson:

1. The distribution of the number of events of any time interval depends only on its length or duration. (eg; number of horse kicks only depends on that it is a day, not on the particular day)

2. The number of events recorded in two disjoint time intervals are independent of one another (number of horse kicks is independent from today and yesterday)
3. No two events are recorded at exactly the same time point (you can't have 2 horsekicks exactly at the same time, must be slightly different in time)

So:

Let $X_{t,s}$ denote the number of events in the time interval $(t, s]$

- Denote by: $X_t = X_{0,t}$

Our goal is to find the distribution of X :

- Let $f(t) = P(X_t = 0)$, then

$$\begin{aligned}
 f(t+s) &= P(X_{t+s} = 0) \\
 &= P(X_t = 0, X_{t,s+t} = 0) \text{ (disjoint time intervals, and using property 2)} \\
 &= P(X_t = 0)P(X_{t,s+t} = 0) \text{ (property 1)} \\
 &= P(X_t = 0)P(X_s = 0) = f(t)f(s) \\
 &\rightarrow f(t+s) = f(t)f(s) \forall t, s > 0
 \end{aligned}$$

One type of solution of this is:

$$f(t) = e^{\alpha t} \quad (\alpha \in \mathbb{R})$$

(note: other solutions exist, but are a mess, in that they are unbounded, but $f(t) \in [0,1]$, so can't exist for probabilities).

For the same reason, $\alpha < 0$ (to remain bounded)

So:

$$\begin{aligned}
 \alpha &= -\lambda \text{ (for } \lambda > 0) \\
 &\rightarrow P(X_t = 0) = e^{-\lambda t} \text{ (for } t = 1) \\
 P_{X_1}(0) &= P(X_1 = 0) = e^{-(\lambda)}
 \end{aligned}$$

Let: Y_n = the number of intervals in $\left(\frac{k-1}{n}, \frac{k}{n}\right]$ in $(0,1]$ where an event occurred is:

$$\begin{aligned}
 &= \sum_{k=1}^n 1_{\left\{X_{\frac{k-1}{n}, \frac{k}{n}} \geq 1\right\}} \text{ (n iid Bernoulli } (p_n) \text{) (as independent of time interval, and disjoint)} \\
 &\therefore p_n = P\left(X_{\frac{k-1}{n}, \frac{k}{n}} \geq 1\right) \\
 &= P\left(X_{\frac{1}{n}} \geq 1\right) \text{ (a least 1)} \\
 &= 1 - P\left(X_{\frac{1}{n}} = 0\right) \text{ (law of total probabilities)}
 \end{aligned}$$

$$= 1 - e^{-\frac{\lambda}{n}}$$

$$\rightarrow \frac{1}{n} \sim \text{Binomial}\left(n, p_n = 1 - e^{-\frac{\lambda}{n}}\right)$$

Note: $\frac{1}{n} \leq X_1$ and $Y_n < X_1$, if two events occur in the same interval

However:

$$\lim_{n \rightarrow \infty} Y_n = X_1$$

(as X_1 accounts ALL events, but Y_n counts only the events in a certain time interval)

Because no two events can occur at the same time. (property 3)

The expected value of a $\text{binomial}(n, p)$ RV is np . For $\frac{1}{n}$, this is $np_n = n\left(1 - e^{-\frac{\lambda}{n}}\right)$, what is the $\lim_{n \rightarrow \infty} np_n$?

$$\begin{aligned} &= n \left(1 - \left(1 - \frac{\lambda}{n} - R_1\left(-\frac{\lambda}{n}\right) \right) \right) \quad (\text{using the Taylor expansion}) \\ &= \lambda + \frac{R_1\left(-\frac{\lambda}{n}\right)}{-\frac{\lambda}{n}} \lambda \rightarrow \lambda \end{aligned}$$

Therefore, the limit of the expected value for the binomial is λ

Claim: if $Y_n \sim \text{Binom}(n, p_n)$ such that $np_n \rightarrow \lambda$ (as $n \rightarrow \infty$), then for any fixed $k \in \mathbb{Z}^+$:

$$P(Y_n = k) \rightarrow_{n \rightarrow \infty} e^{-\lambda} \frac{\lambda^k}{k!}$$

(binomial converges to the poisson pmf)

Corollary:

$$\begin{aligned} X = X_1 &\sim \text{Poisson}(\lambda) \text{ for large } n \\ \lim_{n \rightarrow \infty} P(Y_n = k) &\rightarrow P(X_{\text{poisson}}) \end{aligned}$$

Proof of claim:

$$\begin{aligned} P(Y_n = k) &= \binom{n}{k} p_n^k (1 - p_n)^{n-k} \\ &= \frac{n(n-1) \dots (n-k+1)}{k!} p_n^k \end{aligned}$$

As $n \gg k$:

$$\rightarrow \frac{(np_n)^k}{k!} \left(\frac{n}{n} \frac{n-1}{n} \dots \frac{n-k+1}{n} \right)$$

$$\text{as } n \rightarrow \infty = \frac{\lambda^k}{k!} \quad (\text{using assumption})$$

Also:

$$(1 - p_n)^{n-k} = (1 - p_n)^n (1 - p_n)^k$$

$$\rightarrow (1 - p_n)^n \quad (\text{as } p_n \rightarrow 0 \text{ as } n \rightarrow \infty)$$

$$\therefore \lim_{n \rightarrow \infty} (1 - p_n)^n = e^{n \ln(1 - p_n)}$$

$$\text{Taylor expansion: } \ln(1 - x) = -x + R_1(x), \text{ where } \frac{R(x)}{x} \rightarrow 0$$

$\therefore$ limit is:

$$n \ln(1 - p_n) = n[-p_n + R_1(p_n)]$$

$$= -np_n + \frac{R_1(p_n)}{p_n} (np_n)$$

$$\therefore \text{limit} \rightarrow -\lambda$$

$$\therefore \lim_{n \rightarrow \infty} ((1 - p_n)^n) = e^{-\lambda}$$

Example of poisson approximation to binomial distribution:

$$n = 100; p = 0.01; \lambda = np$$

$$P(Y_n = 3) = 0.0610 \quad Y_n \sim \text{Binom}(n, p)$$

$$P(X_{\text{Poisson}} = 3) \approx 0.0613 \quad (X \sim \text{Poisson}(\lambda))$$

(so pretty good approximation)

Proof of corollary:

$$\lim_{n \rightarrow \infty} P(Y_n = k)$$

Going back to our horse deaths:

What is missing?

λ is missing, how did he come up with his lambda?

Computing the mean of the empirical data

Example: faulty monitor

The radioactive source gives off particles emitted according to $\text{Poisson}(\lambda)$

How do we see if the monitor is faulty, and if it is picking up all the radiation?

Let X be the number of particle REGISTERED in an hour

What is the distribution of X ?

Let N = the number of particles EMITTED in the same time (note: it is unobserved)

For $k \in \mathbb{N}$;

$$P(X = k) = \sum_{n=k}^{\infty} P(X = n|N = n)P(N = n) \quad (\text{using law of total probability})$$

note: $P(X = n|N = n) = 0$ for all $n \leq k$ (it can't see particles which aren't there)

First is a binomial distribution; $\text{Binom}(n, p)$; second is poisson (given)

$$\begin{aligned} \text{so } P(X = k) &= \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} e^{-\lambda} \frac{\lambda^n}{n!} \\ &= e^{-\lambda} \frac{(\lambda p)^k}{k!} \sum_{n=k}^{\infty} \frac{\lambda^{n-k} (1-p)^{n-k}}{(n-k)!} \quad (\text{doing some algebra}) \\ &\quad \text{letting } m = n - k \\ &= e^{-\lambda} \frac{(\lambda p)^k}{k!} \sum_{m=0}^{\infty} \frac{(\lambda(1-p))^m}{(m)!} \\ &= e^{-\lambda} \frac{(\lambda p)^k}{k!} e^{\lambda(1-p)} \quad (\text{Taylor series}) \\ &= e^{-\lambda p} \frac{(\lambda p)^k}{k!} \end{aligned}$$

SO:

$$X \sim \text{Poisson}(\lambda p)$$

This makes sense, that the three counting assumptions for the poisson distribution still hold

Distribution of a sum of random variables: (arithmeticies of RV)

If X and Y are jointly distributed $\mathbb{Z}^+$ valued RV's; then $Z = X + Y$ is also a $\mathbb{Z}^+$ valued RV, and for $n \in \mathbb{Z}^+$:

Pmf of Z :

$$\begin{aligned} P_Z(n) &= P(Z = n) \\ &= \sum_{k=0}^n P(X = k, Y = n - k) \\ &= \sum_{k=0}^n p_{XY}(k, n - k) \end{aligned}$$

If X and Y are independent; then:

$$p_Z(n) = \sum_{k=0}^n p_X(k) p_Y(n - k)$$

Convolution:

p_Z is the **convolution** of p_X and p_Y

$$p_Z = p_X * p_Y$$

Example:

1. $X \sim \text{Binomial}(m, p)$ is independent of $Y \sim \text{Binomial}(n - m, p)$

$$Z = X + Y$$

It is clear that $Z \sim \text{Binomial}(n, p)$

Formally:

$$\begin{aligned} l &= \{0, \dots, n\} \\ P(Z = l) &= \sum_{k=0}^l \binom{m}{k} p^k (1-p)^{m-k} \binom{n-m}{l-k} p^{l-k} (1-p)^{n-m-(l-k)} \\ &= p^l (1-p)^{n-l} \sum_{k=0}^l \binom{m}{k} \binom{n-m}{l-k} \\ &= p^l (1-p)^{n-l} \binom{n}{l} \end{aligned}$$

Proof that:

$$\sum_{k=0}^l \binom{m}{k} \binom{n-m}{l-k} = \binom{n}{l}$$

Combinatorics:

$\binom{n}{l}$ is the number of committees of l students from a class of n

Suppose the class has m boys and $n - m$ girls; then the same number can be counted as

$$\sum_{k=0}^l \{\text{number of committees with } k \text{ boys}\} = \sum_{k=0}^l \binom{m}{k} \binom{n-m}{l-k}$$

Algebraic proof:

$$(1+x)^n = (1+x)^m (1+x)^{n-m}$$

consider the coefficient of x^l on both sides

$$LHS = \binom{n}{l}; \quad RHS = \sum_{k=0}^l \binom{m}{k} \binom{n-m}{l-k}$$

Example 2:

$X \sim \text{Poisson}(\lambda); Y \sim \text{Poisson}(\gamma);$ what is $P(Z = X + Y)$?

$$P(Z = n) = \sum_{k=0}^n p_X(k)p_Y(n-k) = \sum_{k=0}^n \frac{e^{-\lambda}\lambda^k}{k!} \frac{e^{-\gamma}\gamma^{n-k}}{(n-k)!}$$

Example 3

$$X \sim B(n, p); Y \sim B(n-m, p)$$

We already saw $X + Y \sim B(n, p)$, ut now condider the CONDITIONAL distribution of X , given $X + Y$

For $k = 0, \dots, n$

$X \sim \text{Poisson}(\lambda)$ and $Y \sim P(\gamma)$; we say $X + Y \sim Po(\lambda + \gamma)$

The conditional distribution of X is

$$P(X = k | X + Y = n) = \frac{P(X = k, Y = n - k)}{P(X + Y = n)} = \frac{\frac{e^{-\lambda}\lambda^k}{k!} \frac{e^{-\gamma}\gamma^{n-k}}{(n-k)!}}{\frac{e^{-(\lambda+\gamma)}(\lambda+\gamma)^n}{n!}}$$

$$\rightarrow X \sim \text{Binom}\left(n, \frac{\lambda}{\lambda + \gamma}\right)$$

Imagine the sources emitting particles at rates λ and γ

Expectation:

Definition

Def: the expected value, or expectation, of a discrete non-negative RV X is the weighted average with range $X(\Omega) = \{x_i\}$ is given by:

$$E(X) = \sum_i x_i p_i(x_i)$$

Well definied for positive and negative separately; not well defined for if it is positive and negative

Examples of particular kinds

Bernoulli:

$$X \sim \text{Bernoulli}(p)$$

$$E(X) = 0(1-p) + 1(p) = p$$

Poisson

$$X \sim \text{POisson}(\lambda)$$

$$\begin{aligned}
 E(X) &= \sum_k \frac{k e^{-\lambda} \lambda^k}{k!} \\
 &= \lambda e^{-\lambda} e^{\lambda} \\
 &= \lambda
 \end{aligned}$$

Binomial

$$\begin{aligned}
 &= \sum_k k \binom{n}{k} p^k (1-p)^{n-k} \\
 &= np \sum_{k=1}^n \frac{(n-1)!}{(k-1)! [(n-1) - (k-1)]!} p^{k-1} (1-p)^{n-1-(k-1)} \\
 &= np \sum_{m=1}^{n-1} \binom{n-1}{m} p^m (1-p)^{n-1-m} \\
 &= np(p + (1-p)) \\
 &= np
 \end{aligned}$$

Geometric:

$$E(X) = \sum_{k=1}^{\infty} k(1-p)^{k-1}p$$

Consider the power series:

$$f(q) = \sum_{k=0}^m q^k = \frac{1}{1-q} \quad \text{for } |q| < 1$$

A power series can be differentiated term by term within its radius of convergence

$$\begin{aligned}
 \rightarrow f'(q) &= \sum_{k=1}^m k q^{k-1} \quad \forall q \in (-1, 1) \\
 \text{but } f'(q) &= \frac{d}{dq} \left(\frac{1}{1-q} \right) = \frac{1}{(1-q)^2} \\
 \rightarrow \frac{1}{(1-q)^2} &= \sum_{k=1}^m k q^{k-1} \\
 \therefore E(X) &= p \sum_{k=1}^{\infty} \frac{k q^{k-1} p}{(1-q)^2} \\
 &= \frac{1}{p}
 \end{aligned}$$

Negative binomial

$$E(X) = \frac{r}{p}$$